

A smoothing-based conjugate gradient approach for image restoration

Nurullah Yilmaz^{†*}, Ahmet Sahiner[‡], Gulden Kapusuz Gokdogan[‡]

[†]*Department of Mathematics, Suleyman Demirel University, Isparta, Turkiye*

[‡]*Department of Basic Sciences, Isparta University of Applied Sciences, Isparta, Turkiye*

Email(s): nurullahyilmaz@sdu.edu.tr, ahmetsahiner@sdu.edu.tr, d_guldenkapusuz@isparta.edu.tr

Abstract. This paper studies image restoration using Total Variation, non-convex Lipschitz, and non-Lipschitz regularizers. We use indicator-based smoothing functions (local polynomial and global arc-tangent) within a Smoothing Conjugate Gradient Algorithm (SCGA) and an inner-outer continuation framework. The search direction satisfies the sufficient descent condition. For locally Lipschitz models, global convergence to Clarke stationary points is established. For the non-Lipschitz model, the limit points are shown to satisfy asymptotic ε -stationarity. Numerical experiments over independent noise realizations evaluate the proposed variants against baseline solvers. The results indicate that the non-convex Lipschitz and non-Lipschitz models obtain higher PSNR under heavy noise, whereas baseline methods maintain competitive SSIM and CPU times in impulse noise settings.

Keywords: Smoothing techniques; conjugate gradient method; non-smooth optimization; image restoration.

AMS Subject Classification 2010: 65D10, 90C26, 65K10.

1 Introduction

Image restoration plays a fundamental role in scientific and industrial applications, including defense systems, medical imaging, astronomy, and microscopy. Due to physical limitations in acquisition devices, captured images are inevitably degraded by blur and noise, requiring effective numerical techniques to reconstruct an underlying true image from a degraded observation prior to downstream analysis [15,29,30]. To address this inverse problem, we focus on the general unconstrained optimization formulation:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \|Ax - b\|^2 + \lambda \sum_{i=1}^r \varphi(d_i^T x), \quad (1)$$

where $A \in \mathbb{R}^{m \times n}$ denotes the blurring matrix, $b \in \mathbb{R}^m$ is the observed image vector, $\lambda > 0$ is a regularization parameter, and $d_i^T x$ computes the first-order discrete differences between adjacent pixels

*Corresponding author

Received: 06 October 2026/ Revised: 22 October 2026/ Accepted: 24 October 2026

DOI: [10.22124/jmm.2026.35108.3280](https://doi.org/10.22124/jmm.2026.35108.3280)

for $i = 1, 2, \dots, r$, with d_i^T representing the i -th row of the difference matrix $D \in \mathbb{R}^{r \times n}$. In (1), the first term ensures data fidelity, while the second serves as a regularizer incorporating prior structural knowledge. Since image restoration is inherently ill-posed and formulations involving non-convex regularizers often lead to NP-hard problems, an appropriate prior is essential to narrow down the feasible solution space [30]. Various potential functions $\varphi(t)$ have been investigated in the literature. Classical Tikhonov regularization ($\varphi(t) = t^2$) is smooth and convex [15], allowing fast transform-based computation, but suffers from severe edge-blurring [25]. Conversely, Total Variation (TV) regularization, defined as $\varphi_1(t) = |t|$, is non-smooth and convex [28, 33, 34], effectively preserving sharp boundaries and providing robustness against outliers [21]. To further enhance contrast and recover piecewise constant regions with sharper object contours [18, 27, 42], non-smooth non-convex regularizers have been introduced, expressed either via the Lipschitz continuous potential $\varphi_2(t) = \frac{\alpha|t|}{1+\alpha|t|}$ ($\alpha > 0$) [26] or the non-Lipschitz potential $\varphi_3(t) = |t|^p$ ($0 < p < 1$) [5, 48]. Consequently, we analyze and evaluate the models based on $\varphi_1(t)$, $\varphi_2(t)$, and $\varphi_3(t)$ separately.

Numerous optimization algorithms have been developed for non-smooth ℓ_1 -norm and TV regularized inverse problems, including iterative shrinkage-thresholding [3, 12], variable splitting [1], NESTA [4], fixed-point continuation [16], and Bregman iterative methods [47]. While effective in convex settings, extending them to non-convex or non-Lipschitz models often requires solving demanding subproblems at each iteration. To overcome these bottlenecks in large-scale settings, non-linear Conjugate Gradient (CG) methods—originating from Hestenes and Stiefel [17] and further developed by Fletcher and Reeves [13, 14], Polak and Ribière [31], Polyak [32], Liu and Storey [22], and Dai and Yuan [11]—have gained attention for their low memory footprint and efficiency, inspiring modified and three-term CG variants tailored for image restoration [6, 19, 20, 37]. However, standard CG solvers cannot be directly applied to (1) due to the non-differentiability of $|\cdot|$ and $|\cdot|^p$ at the origin [9, 26, 27]. A widely adopted remedy is numerical smoothing [2, 7, 10, 24, 40, 45, 46], which approximates the non-smooth objective via a parameterized family of continuously differentiable functions [23, 35, 36]. Building on the pioneering smoothing CG (SCG) framework of Chen and Zhou [8] and the smoothing PRP scheme of Zhu [49], recent studies have combined CG solvers with various smoothing approximations for ℓ_1 -regularized signal recovery [38, 39]. Nevertheless, a unified smoothing-CG framework capable of seamlessly handling convex, non-convex Lipschitz, and non-Lipschitz regularizers in image restoration remains lacking.

To address this, we construct indicator-based local polynomial and global arctangent smoothing functions that uniformly approximate convex (φ_1), non-convex Lipschitz (φ_2), and non-Lipschitz (φ_3) regularizers. These approximations are integrated into a Smoothing Conjugate Gradient Algorithm (SCGA) within an inner-outer continuation framework. We show that the search direction satisfies the sufficient descent condition. Global convergence to Clarke stationary points is established for the locally Lipschitz models, and asymptotic ε -stationarity is shown for the non-Lipschitz case. Numerical experiments on benchmark images evaluate the proposed approach against baseline solvers. The remainder of this paper is organized as follows: Section 2 reviews preliminary definitions. Section 3 formulates the smoothing functions and error bounds. Section 4 details the SCGA framework and its convergence. Section 5 presents the numerical experiments, and finally Section 6 concludes the paper.

2 Preliminaries

Throughout this paper, $\mathbb{R}^n$ denotes the n -dimensional Euclidean space equipped with the standard inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$, and $\mathbb{R}_{++} = \{t \in \mathbb{R} : t > 0\}$. Because the potential functions in (1) are non-

differentiable at the origin, standard gradients cannot be directly utilized. By Rademacher's Theorem, locally Lipschitz continuous functions (such as φ_1 and φ_2) are differentiable almost everywhere, allowing the use of Clarke's generalized gradient.

Definition 1. ([8]) Let $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be a locally Lipschitz continuous function, and let $D_h \subset \mathbb{R}^n$ denote the set of points where h is differentiable. The Clarke generalized gradient of h at $x \in \mathbb{R}^n$ is defined as:

$$\partial h(x) = \text{conv} \left\{ \lim_{k \rightarrow \infty} \nabla h(x_k) : x_k \rightarrow x, x_k \in D_h \right\}, \quad (2)$$

where conv denotes the convex hull. A point $x^* \in \mathbb{R}^n$ is called a Clarke stationary point of h if $0 \in \partial h(x^*)$. While Clarke stationarity is well-defined for the locally Lipschitz regularizers $\varphi_1(t)$ and $\varphi_2(t)$, the non-Lipschitz potential $\varphi_3(t) = |t|^p$ ($0 < p < 1$) is not locally Lipschitz at $t = 0$. Consequently, our Clarke-based global convergence guarantees are strictly restricted to the locally Lipschitz models ($l \in \{1, 2\}$), whereas stationarity for the non-Lipschitz model ($l = 3$) is understood in the sense of the limiting (Mordukhovich) subdifferential $0 \in \partial_L f(x^*)$ or asymptotic ε -stationarity [5].

To overcome non-smoothness in gradient-based optimization, we approximate the non-smooth terms via parameterized smoothing functions.

Definition 2. ([5, 39]) Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function. A function $\tilde{h} : \mathbb{R} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ is called a smoothing function of h if, for any fixed smoothing parameter $\varepsilon > 0$, $\tilde{h}(\cdot, \varepsilon)$ is continuously differentiable on $\mathbb{R}$, $\tilde{h}(t, \cdot)$ is differentiable on $\mathbb{R}_{++}$ for any fixed $t \in \mathbb{R}$, and $\lim_{\varepsilon \downarrow 0} \tilde{h}(t, \varepsilon) = h(t)$. Moreover, if $h(\cdot)$ is locally Lipschitz continuous near t , it satisfies the gradient consistency property:

$$\left\{ \lim_{z \rightarrow t, \varepsilon \downarrow 0} \nabla_z \tilde{h}(z, \varepsilon) \right\} \subseteq \partial h(t). \quad (3)$$

3 New generation smoothing techniques

This section constructs smoothing approximations for the nonsmooth potential functions $\varphi_1(t) = |t|$, $\varphi_2(t) = \frac{\alpha|t|}{1+\alpha|t|}$, and $\varphi_3(t) = |t|^p$ ($0 < p < 1$). To position our contributions relative to existing literature, we note that classical restoration methods typically employ hyperbolic smoothing ($\sqrt{t^2 + \varepsilon^2}$) or Huber-type approximations, which approach $|t|$ relatively slowly or complicate non-Lipschitz extensions. Building upon indicator-based local polynomial smoothing previously studied for absolute value equations [40, 41, 43], the present work introduces three specific advances: (i) a globally defined arctangent smoothing function $\tilde{\rho}_2(t, \varepsilon)$ providing a single analytical expression on $\mathbb{R}$ with well-behaved gradients for non-convex optimization; (ii) C^1 and C^2 polynomial transition functions ($\tilde{\mu}_1, \tilde{\mu}_2$) tailored specifically for the non-Lipschitz regularizer $\mu(t) = |t|^p$ ($0 < p < 1$); and (iii) explicit uniform $O(\varepsilon)$ and $O(\varepsilon^p)$ approximation error bounds depending solely on ε (and p) without increasing the algebraic cost of Conjugate Gradient iterations. In light of Definition 2, a function $\tilde{s}(t, \varepsilon)$ is a valid smoothing approximation of $s(t) = |t|^p$ ($0 < p \leq 1$) if it is C^1 -differentiable for any fixed $\varepsilon > 0$ and satisfies $\lim_{\varepsilon \rightarrow 0^+} \tilde{s}(t, \varepsilon) = |t|^p$. Henceforth, we denote $\rho(t) = |t|$ and $\mu(t) = |t|^p$ ($0 < p < 1$).

3.1 Smoothing functions for $\rho(t) = |t|$

Following the indicator-based reformulation in [40, 41, 43], $\rho(t) = |t|$ is expressed as:

$$|t| = t\chi_A(t) - t\chi_{A^c}(t) = t(2\chi_A(t) - 1), \quad (4)$$

where $A = [0, \infty)$, $\chi_A : \mathbb{R} \rightarrow \{0, 1\}$ is the indicator function, and $\chi_{A^c}(t) = 1 - \chi_A(t)$. Since non-differentiability at $t = 0$ stems solely from the step discontinuity of $\chi_A(t)$, our first approach replaces $\chi_A(t)$ with a local polynomial smoothing function $\tilde{\chi}_A^1(t, \varepsilon)$ on $\mathbb{R} \times \mathbb{R}_{++}$ [40]:

$$\tilde{\chi}_A^1(t, \varepsilon) = \begin{cases} 0, & t \leq -\varepsilon, \\ P(t, \varepsilon), & -\varepsilon < t < \varepsilon, \\ 1, & t \geq \varepsilon, \end{cases} \quad (5)$$

where $\varepsilon > 0$ and $P(t, \varepsilon)$ is a polynomial transition matching the boundary conditions at $t = \pm\varepsilon$. To achieve first-order (C^1) and second-order (C^2) smoothness, we employ the transition polynomials $P_1(t, \varepsilon)$ and $P_2(t, \varepsilon)$, respectively:

$$P_1(t, \varepsilon) = -\frac{1}{4\varepsilon^3}t^3 + \frac{3}{4\varepsilon}t + \frac{1}{2}, \quad P_2(t, \varepsilon) = \frac{3}{16\varepsilon^5}t^5 - \frac{10}{16\varepsilon^3}t^3 + \frac{15}{16\varepsilon}t + \frac{1}{2}. \quad (6)$$

Substituting $P_1(t, \varepsilon)$ and $P_2(t, \varepsilon)$ into (4) yields the local C^1 and C^2 smoothed approximations $\tilde{\rho}_1^{(1)}(t, \varepsilon)$ and $\tilde{\rho}_1^{(2)}(t, \varepsilon)$:

$$\tilde{\rho}_1^{(1)}(t, \varepsilon) = \begin{cases} |t|, & |t| \geq \varepsilon, \\ -\frac{1}{2\varepsilon^3}t^4 + \frac{3}{2\varepsilon}t^2, & |t| < \varepsilon, \end{cases} \quad \tilde{\rho}_1^{(2)}(t, \varepsilon) = \begin{cases} |t|, & |t| \geq \varepsilon, \\ \frac{3}{8\varepsilon^5}t^6 - \frac{5}{4\varepsilon^3}t^4 + \frac{15}{8\varepsilon}t^2, & |t| < \varepsilon. \end{cases} \quad (7)$$

Our second approach utilizes a globally defined arctangent indicator approximation $\tilde{\chi}_A^2(t, \varepsilon)$, which upon substitution into (4) yields the globally smoothed function $\tilde{\rho}_2(t, \varepsilon)$:

$$\tilde{\chi}_A^2(t, \varepsilon) = \frac{1}{\pi} \arctan\left(\frac{t}{\varepsilon}\right) + \frac{1}{2}, \quad \tilde{\rho}_2(t, \varepsilon) = t(2\tilde{\chi}_A^2(t, \varepsilon) - 1) = \frac{2t}{\pi} \arctan\left(\frac{t}{\varepsilon}\right), \quad (8)$$

where $\lim_{t \rightarrow \infty} \tilde{\chi}_A^2(t, \varepsilon) = 1$, $\lim_{t \rightarrow -\infty} \tilde{\chi}_A^2(t, \varepsilon) = 0$, and $\tilde{\chi}_A^2(0, \varepsilon) = \frac{1}{2}$.

Lemma 1. ([40]) Let $\chi_A(t)$ be the indicator function of $A = [0, \infty) \subset \mathbb{R}$, and let $\tilde{\chi}_A^j(t, \varepsilon)$ denote its smooth approximation for $j = 1, 2$. Then, for any $\varepsilon > 0$ and $t \in \mathbb{R}$, $|\tilde{\chi}_A^j(t, \varepsilon) - \chi_A(t)| \leq \frac{1}{2}$.

Unifying both constructions, we write $\tilde{\rho}_j(t, \varepsilon) = 2t\tilde{\chi}_A^j(t, \varepsilon) - t$ for $j = 1, 2$, where $j = 1$ denotes the local polynomial approximation $\tilde{\rho}_1(t, \varepsilon) \in \{\tilde{\rho}_1^{(1)}(t, \varepsilon), \tilde{\rho}_1^{(2)}(t, \varepsilon)\}$ and $j = 2$ denotes the global arctangent approximation $\tilde{\rho}_2(t, \varepsilon)$.

Lemma 2. Let $t \in \mathbb{R}$ and $\varepsilon > 0$. Then: (i) $\tilde{\rho}_1^{(1)}(t, \varepsilon)$ and $\tilde{\rho}_2(t, \varepsilon)$ are C^1 -continuously differentiable on $\mathbb{R}$ (while $\tilde{\rho}_1^{(2)}(t, \varepsilon)$ is C^2 on $\mathbb{R}$); (ii) the approximation errors scale linearly with ε and satisfy the exact uniform bounds:

$$|\tilde{\rho}_1^{(1)}(t, \varepsilon) - |t|| \leq \frac{3\sqrt{3}-5}{4}\varepsilon \leq \frac{\varepsilon}{2}, \quad |\tilde{\rho}_1^{(2)}(t, \varepsilon) - |t|| \leq \frac{3}{8}\varepsilon \leq \frac{\varepsilon}{2}, \quad |\tilde{\rho}_2(t, \varepsilon) - |t|| < \frac{2}{\pi}\varepsilon. \quad (9)$$

Consequently, $|\tilde{\rho}_j(t, \varepsilon) - |t|| \leq C_j\varepsilon$ holds for all $t \in \mathbb{R}$, where $C_1 = \frac{3\sqrt{3}-5}{4} \approx 0.1745$ for $\tilde{\rho}_1^{(1)}$ (or $C_1 = \frac{3}{8} = 0.375$ for $\tilde{\rho}_1^{(2)}$, and unified as $C_1 \leq \frac{1}{2}$ for both local variants) and $C_2 = \frac{2}{\pi} \approx 0.6366$ for $\tilde{\rho}_2$, with the overall bound $|\tilde{\rho}_j(t, \varepsilon) - |t|| \leq C\varepsilon$ where $C = \max\{C_1, C_2\} = \frac{2}{\pi}$; and (iii) $\lim_{\varepsilon \rightarrow 0^+} \tilde{\rho}_j(t, \varepsilon) = |t|$ uniformly on $\mathbb{R}$.

Proof. (i) Differentiating $\tilde{\rho}_j(t, \varepsilon)$ with respect to t yields $\tilde{\rho}_j'(t, \varepsilon) = 2\tilde{\chi}_A^j(t, \varepsilon) + 2t(\tilde{\chi}_A^j)'(t, \varepsilon) - 1$. For local smoothing ($j = 1$), $(\tilde{\chi}_A^1)'(t, \varepsilon) = P'(t, \varepsilon)$ for $|t| < \varepsilon$ and 0 for $|t| \geq \varepsilon$, where $P_1'(t, \varepsilon) = -\frac{3}{4\varepsilon^3}t^2 + \frac{3}{4\varepsilon}$ and $P_2'(t, \varepsilon) = \frac{15}{16\varepsilon^5}t^4 - \frac{30}{16\varepsilon^3}t^2 + \frac{15}{16\varepsilon}$. At $t = \pm\varepsilon$, $\lim_{t \rightarrow \pm\varepsilon}(\tilde{\rho}_1^{(m)})'(t, \varepsilon) = \pm 1$ for $m = 1, 2$ (and $\lim_{t \rightarrow \pm\varepsilon}(\tilde{\rho}_1^{(2)})''(t, \varepsilon) = 0$), ensuring C^1 - and C^2 -continuity, respectively. For global smoothing ($j = 2$), $(\tilde{\chi}_A^2)'(t, \varepsilon) = \frac{\varepsilon}{\pi(\varepsilon^2 + t^2)}$, which is smooth everywhere on $\mathbb{R}$.

(ii) For $j = 1$, $|\tilde{\rho}_1^{(m)}(t, \varepsilon) - |t|| = 0$ when $|t| \geq \varepsilon$ ($m = 1, 2$). For $|t| < \varepsilon$, by even symmetry we restrict attention to $t \in [0, \varepsilon)$ and introduce the dimensionless variable $s = t/\varepsilon \in [0, 1)$, so $t = \varepsilon s$. For $m = 1$, substituting $t = \varepsilon s$ into (7) shows that the error scales linearly with ε :

$$E_1^{(1)}(t) = \tilde{\rho}_1^{(1)}(\varepsilon s, \varepsilon) - \varepsilon s = \varepsilon \left(-\frac{1}{2}s^4 + \frac{3}{2}s^2 - s \right) =: \varepsilon \hat{e}_1(s). \quad (10)$$

Differentiating with respect to s gives $\hat{e}_1'(s) = -2s^3 + 3s - 1 = -(s-1)(2s^2 + 2s - 1) = 0$, which has a unique root in $(0, 1)$ at $s^* = \frac{\sqrt{3}-1}{2} \approx 0.3660$. Evaluating $|\hat{e}_1(s^*)|$ yields $\max_{t \in \mathbb{R}} |E_1^{(1)}(t)| = \frac{3\sqrt{3}-5}{4}\varepsilon \approx 0.1745\varepsilon \leq \frac{\varepsilon}{2}$. Similarly, for $m = 2$, setting $t = \varepsilon s$ yields $E_1^{(2)}(t) = \varepsilon \left(\frac{3}{8}s^6 - \frac{5}{4}s^4 + \frac{15}{8}s^2 - s \right)$, which satisfies $\max_{s \in [0, 1]} |E_1^{(2)}(\varepsilon s)| \leq \frac{3}{8}\varepsilon \leq \frac{\varepsilon}{2}$. For $j = 2$, $E_2(t) = \frac{2t}{\pi} \arctan(\frac{t}{\varepsilon}) - |t|$ is even, and for $t \geq 0$, $E_2(t) = -\frac{2t}{\pi} \operatorname{arccot}(\frac{t}{\varepsilon})$. Using $\operatorname{arccot}(y) < \frac{1}{y}$ for $y = \frac{t}{\varepsilon} > 0$ gives $|E_2(t)| = \frac{2\varepsilon}{\pi} \operatorname{arccot}(y) < \frac{2}{\pi}\varepsilon$. (iii) Taking $\varepsilon \rightarrow 0^+$ in (9) establishes uniform convergence on $\mathbb{R}$. $\square$

3.2 Smoothing functions for $\mu(t) = |t|^p$ ($0 < p < 1$)

Following the piecewise local smoothing strategy in [44], we approximate the non-convex, non-Lipschitz potential $\mu(t) = |t|^p$ ($0 < p < 1$) via:

$$\tilde{\mu}_i(t, \varepsilon) = \begin{cases} |t|^p, & |t| > \varepsilon, \\ \eta_i(t, \varepsilon), & |t| \leq \varepsilon, \end{cases} \quad (i = 1, 2), \quad (11)$$

where $\eta_1(t, \varepsilon)$ and $\eta_2(t, \varepsilon)$ are C^1 and C^2 polynomial transition functions matching the boundary values and derivatives at $t = \pm\varepsilon$:

$$\eta_1(t, \varepsilon) = \frac{p}{2}\varepsilon^{p-2}t^2 + \frac{2-p}{2}\varepsilon^p, \quad \eta_2(t, \varepsilon) = \frac{p(p-2)}{8}\varepsilon^{p-4}t^4 - \frac{p(p-4)}{4}\varepsilon^{p-2}t^2 + \frac{(p-2)(p-4)}{8}\varepsilon^p. \quad (12)$$

Lemma 3. Let $\mu(t) = |t|^p$ ($0 < p < 1$) and $\tilde{\mu}_i(t, \varepsilon)$ be defined by (11)–(12). Then: (i) $\tilde{\mu}_1(t, \varepsilon)$ is C^1 and $\tilde{\mu}_2(t, \varepsilon)$ is C^2 with respect to t on $\mathbb{R}$; (ii) for all $t \in \mathbb{R}$ and $\varepsilon > 0$, both approximations majorize $|t|^p$ ($\tilde{\mu}_i(t, \varepsilon) - |t|^p \geq 0$), and the approximation errors satisfy:

$$0 \leq \tilde{\mu}_1(t, \varepsilon) - |t|^p \leq \frac{2-p}{2}\varepsilon^p, \quad \text{and} \quad 0 \leq \tilde{\mu}_2(t, \varepsilon) - |t|^p \leq \frac{(p-2)(p-4)}{8}\varepsilon^p; \quad (13)$$

and (iii) $\lim_{\varepsilon \rightarrow 0^+} \tilde{\mu}_i(t, \varepsilon) = |t|^p$ uniformly on $\mathbb{R}$ for $i = 1, 2$.

Proof. (i) Smoothness holds trivially for $|t| \neq \varepsilon$. At $t = \varepsilon$, direct evaluation gives $\eta_i(\varepsilon, \varepsilon) = \varepsilon^p = \mu(\varepsilon)$ and $\eta_i'(\varepsilon, \varepsilon) = p\varepsilon^{p-1} = \mu'(\varepsilon)$ for $i = 1, 2$, while $\eta_2''(\varepsilon, \varepsilon) = p(p-1)\varepsilon^{p-2} = \mu''(\varepsilon)$; even symmetry ensures matching at $t = -\varepsilon$. (ii) Define the error functions $E_i(t) := \tilde{\mu}_i(t, \varepsilon) - |t|^p$ for $i = 1, 2$. Since

$E_i(t) = 0$ for $|t| \geq \varepsilon$ and $E_i(-t) = E_i(t)$, it suffices to analyze $E_i(t)$ on $[0, \varepsilon]$. For $i = 1$ and $t \in (0, \varepsilon)$, differentiating $E_1(t) = \eta_1(t, \varepsilon) - t^p$ yields $E_1'(t) = pt(\varepsilon^{p-2} - t^{p-2})$. Since $0 < p < 1$ implies $2 - p > 0$, we have $t^{p-2} > \varepsilon^{p-2}$ for all $t \in (0, \varepsilon)$, so the derivative is strictly negative ($E_1'(t) < 0$). Consequently, $E_1(t)$ is strictly decreasing on $[0, \varepsilon]$ from its maximum $E_1(0) = \eta_1(0, \varepsilon) = \frac{2-p}{2}\varepsilon^p > 0$ down to its minimum $E_1(\varepsilon) = 0$, which proves that $E_1(t) \geq 0$ and establishes the first bound in (13). For $i = 2$ and $t \in (0, \varepsilon)$, differentiating $E_2(t) = \eta_2(t, \varepsilon) - t^p$ gives:

$$E_2'(t) = \frac{p(p-2)}{2}\varepsilon^{p-4}t^3 - \frac{p(p-4)}{2}\varepsilon^{p-2}t - pt^{p-1} = pt\varepsilon^{p-2} \left[\frac{p-2}{2} \left(\frac{t}{\varepsilon} \right)^2 + \frac{4-p}{2} - \left(\frac{t}{\varepsilon} \right)^{p-2} \right]. \quad (14)$$

Letting $s = t/\varepsilon \in (0, 1)$ and $\phi(s) := \frac{p-2}{2}s^2 + \frac{4-p}{2} - s^{p-2}$, we observe that $\phi(1) = 0$ and $\phi'(s) = (p-2)s - (p-2)s^{p-3} = (2-p)s(s^{p-4} - 1) > 0$ for all $s \in (0, 1)$ because $0 < p < 1$ and $s^{p-4} > 1$. Since $\phi(s)$ is strictly increasing on $(0, 1)$ with $\phi(1) = 0$, we have $\phi(s) < 0$ for all $s \in (0, 1)$, which implies $E_2'(t) < 0$ on $(0, \varepsilon)$. Thus, the error function $E_2(t)$ is also strictly decreasing on $[0, \varepsilon]$, attaining its minimum value $\min_{t \in [0, \varepsilon]} E_2(t) = E_2(\varepsilon) = 0$ (hence $E_2(t) \geq 0$ is non-negative everywhere) and its maximum value at $t = 0$:

$$\max_{t \in [0, \varepsilon]} E_2(t) = E_2(0) = \eta_2(0, \varepsilon) = \frac{(p-2)(p-4)}{8}\varepsilon^p > 0. \quad (15)$$

(iii) Taking $\varepsilon \rightarrow 0^+$ in (13) yields uniform convergence since $p > 0$. $\square$

3.3 Smoothed problem formulation and error analysis

Using $\tilde{\rho}_j(t, \varepsilon)$ ($j = 1, 2$) for φ_1, φ_2 and $\tilde{\mu}_i(t, \varepsilon)$ ($i = 1, 2$) for φ_3 , the smoothed optimization problem associated with (1) is formulated as:

$$\min_{x \in \mathbb{R}^n} \tilde{f}_l(x, \varepsilon) = \frac{1}{2} \|Ax - b\|^2 + \lambda \tilde{\Phi}_l(Dx, \varepsilon), \quad \text{with} \quad \tilde{\Phi}_l(Dx, \varepsilon) = \sum_{i=1}^r \tilde{\varphi}_l(d_i^T x, \varepsilon), \quad (l = 1, 2, 3), \quad (16)$$

where $\lambda > 0$, $D \in \mathbb{R}^{r \times n}$ has rows d_i^T , and the smoothed potential functions are given by:

$$\tilde{\varphi}_1(t, \varepsilon) = \tilde{\rho}_j(t, \varepsilon), \quad \tilde{\varphi}_2(t, \varepsilon) = \frac{\alpha \tilde{\rho}_j(t, \varepsilon)}{1 + \alpha \tilde{\rho}_j(t, \varepsilon)} \quad (\alpha > 0), \quad \tilde{\varphi}_3(t, \varepsilon) = \tilde{\mu}_i(t, \varepsilon) \quad (0 < p < 1). \quad (17)$$

Theorem 1. Let $x \in \mathbb{R}^n$ and $\varepsilon > 0$. For any $l \in \{1, 2, 3\}$, there exists $K_l > 0$ such that:

$$|f(x) - \tilde{f}_l(x, \varepsilon)| \leq r\lambda K_l \psi_l(\varepsilon), \quad (18)$$

where $\psi_l(\varepsilon) = \varepsilon$ for $l = 1, 2$ and $\psi_3(\varepsilon) = \varepsilon^p$. Specifically, $K_1 = C_j$ and $K_2 = \alpha C_j$, where $C_1 = \frac{3\sqrt{3}-5}{4} \leq \frac{1}{2}$ for C^1 local smoothing $\tilde{\rho}_1^{(1)}$ (or $C_1 = \frac{3}{8} \leq \frac{1}{2}$ for C^2 local smoothing $\tilde{\rho}_1^{(2)}$) and $C_2 = \frac{2}{\pi}$ for global smoothing $\tilde{\rho}_2$; and for $l = 3$, $K_3 = \frac{2-p}{2}$ (for $i = 1$) or $K_3 = \frac{(p-2)(p-4)}{8}$ (for $i = 2$).

Proof. Canceling the identical fidelity terms gives $|f(x) - \tilde{f}_l(x, \varepsilon)| \leq \lambda \sum_{i=1}^r |\varphi_l(d_i^T x) - \tilde{\varphi}_l(d_i^T x, \varepsilon)|$. Applying Lemma 2 yields $|\varphi_1(t) - \tilde{\varphi}_1(t, \varepsilon)| \leq C_j \varepsilon$ for $l = 1$, and for $l = 2$, since $(1 + \alpha|t|)(1 + \alpha \tilde{\rho}_j(t, \varepsilon)) \geq 1$, we have $|\varphi_2(t) - \tilde{\varphi}_2(t, \varepsilon)| = \frac{\alpha|t| - \tilde{\rho}_j(t, \varepsilon)}{(1 + \alpha|t|)(1 + \alpha \tilde{\rho}_j(t, \varepsilon))} \leq \alpha C_j \varepsilon$. For $l = 3$, Lemma 3 gives $|\varphi_3(t) - \tilde{\varphi}_3(t, \varepsilon)| \leq K_3 \varepsilon^p$. Summing over $i = 1, \dots, r$ completes the proof. $\square$

To apply gradient-based methods, the gradient of $\tilde{f}_l(x, \varepsilon)$ in (16) is written as:

$$g(x, \varepsilon) := \nabla \tilde{f}_l(x, \varepsilon) = A^T(Ax - b) + \lambda D^T \nabla \tilde{\Phi}_l(Dx, \varepsilon), \quad (19)$$

where $\nabla \tilde{\Phi}_l(Dx, \varepsilon) = (\tilde{\Phi}'_l(d_1^T x, \varepsilon), \dots, \tilde{\Phi}'_l(d_r^T x, \varepsilon))^T \in \mathbb{R}^r$.

Lemma 4. For any fixed $\varepsilon > 0$ and $l \in \{1, 2, 3\}$, $\nabla \tilde{f}_l(\cdot, \varepsilon)$ is Lipschitz continuous on $\mathbb{R}^n$ with constant $L_l(\varepsilon) = \|A^T A\| + \lambda \|D\|^2 M_l(\varepsilon)$, where $M_l(\varepsilon) = \sup_{t \in \mathbb{R} \setminus \{\pm \varepsilon\}} |\tilde{\Phi}''_l(t, \varepsilon)| < \infty$; that is, $\|\nabla \tilde{f}_l(x, \varepsilon) - \nabla \tilde{f}_l(y, \varepsilon)\| \leq L_l(\varepsilon) \|x - y\|$ for all $x, y \in \mathbb{R}^n$.

Proof. From (19) and the fundamental theorem of calculus for piecewise C^2 functions, $\|\nabla \tilde{f}_l(x, \varepsilon) - \nabla \tilde{f}_l(y, \varepsilon)\| \leq \|A^T A(x - y)\| + \lambda \|D\| \|\nabla \tilde{\Phi}_l(Dx, \varepsilon) - \nabla \tilde{\Phi}_l(Dy, \varepsilon)\| \leq (\|A^T A\| + \lambda \|D\|^2 M_l(\varepsilon)) \|x - y\|$, where $M_l(\varepsilon) < \infty$ because $\tilde{\Phi}'_l(\cdot, \varepsilon)$ is continuous on $\mathbb{R}$ and has bounded piecewise continuous derivatives on $\mathbb{R} \setminus \{\pm \varepsilon\}$ for any fixed $\varepsilon > 0$. $\square$

Lemma 5. For any fixed $\varepsilon > 0$, initial point $x_0 \in \mathbb{R}^n$, and assuming A has full column rank ($\ker(A) = \{0\}$), the level set $\mathcal{L}(x_0, \varepsilon) = \{x \in \mathbb{R}^n : \tilde{f}_l(x, \varepsilon) \leq \tilde{f}_l(x_0, \varepsilon)\}$ is bounded for $l \in \{1, 2, 3\}$.

Proof. Suppose $\{x_k\} \subset \mathcal{L}(x_0, \varepsilon)$ satisfies $\|x_k\| \rightarrow \infty$. Since $\ker(A) = \{0\}$, A is injective, so $\|Ax_k\| \rightarrow \infty$. Because $\tilde{\Phi}_l(Dx, \varepsilon) \geq 0$, $\tilde{f}_l(x_k, \varepsilon) \geq \frac{1}{2} \|Ax_k - b\|^2 \rightarrow \infty$, contradicting $\tilde{f}_l(x_k, \varepsilon) \leq \tilde{f}_l(x_0, \varepsilon)$. $\square$

Theorem 2. Let $\{\varepsilon_k\}_{k=1}^\infty$ satisfy $\varepsilon_k > 0$ and $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Under the injectivity assumption $\ker(A) = \{0\}$, the coercivity guarantees the existence of global minimizers $x_k^* \in \arg \min_{x \in \mathbb{R}^n} \tilde{f}_l(x, \varepsilon_k)$. Any accumulation point $\bar{x}$ of $\{x_k^*\}$ is a global minimizer of the original problem (1).

Proof. Let $x_{k_j}^* \rightarrow \bar{x}$ as $j \rightarrow \infty$. For any $x \in \mathbb{R}^n$, optimality of $x_{k_j}^*$ and Theorem 1 (with $M_l = r\lambda K_l$) yield:

$$f(x_{k_j}^*) \leq \tilde{f}_l(x_{k_j}^*, \varepsilon_{k_j}) + M_l \psi_l(\varepsilon_{k_j}) \leq \tilde{f}_l(x, \varepsilon_{k_j}) + M_l \psi_l(\varepsilon_{k_j}) \leq f(x) + 2M_l \psi_l(\varepsilon_{k_j}). \quad (20)$$

Taking $j \rightarrow \infty$ and using the continuity of f along with $\psi_l(\varepsilon_{k_j}) \rightarrow 0$ gives $f(\bar{x}) \leq f(x)$ for all $x \in \mathbb{R}^n$. $\square$

Theorem 3. Let $\varepsilon_k \rightarrow 0^+$ and $x_k \rightarrow \bar{x}$ in $\mathbb{R}^n$. For the locally Lipschitz models ($l \in \{1, 2\}$), we have $\lim_{k \rightarrow \infty} \text{dist}(\nabla \tilde{f}_l(x_k, \varepsilon_k), \partial f(\bar{x})) = 0$. In particular, if $\lim_{k \rightarrow \infty} \|\nabla \tilde{f}_l(x_k, \varepsilon_k)\| = 0$, then $0 \in \partial f(\bar{x})$, so $\bar{x}$ is a Clarke stationary point of (1).

Proof. For $l \in \{1, 2\}$, comparing $\nabla \tilde{f}_l(x_k, \varepsilon_k) = A^T(Ax_k - b) + \lambda \sum_{i=1}^r d_i \tilde{\Phi}'_l(d_i^T x_k, \varepsilon_k)$ with the Clarke sub-differential $\partial f(\bar{x}) = A^T(A\bar{x} - b) + \lambda \sum_{i=1}^r d_i \partial \phi_l(d_i^T \bar{x})$ yields:

$$\text{dist}(\nabla \tilde{f}_l(x_k, \varepsilon_k), \partial f(\bar{x})) \leq \|A^T A(x_k - \bar{x})\| + \lambda \sum_{i=1}^r \|d_i\| \text{dist}(\tilde{\Phi}'_l(d_i^T x_k, \varepsilon_k), \partial \phi_l(d_i^T \bar{x})). \quad (21)$$

As $k \rightarrow \infty$, $x_k \rightarrow \bar{x}$ and $\varepsilon_k \rightarrow 0^+$, so both terms on the right-hand side vanish by the continuity of $A^T A$ and the gradient consistency of $\tilde{\Phi}_l$ ($l = 1, 2$). Since $\partial f(\bar{x})$ is closed and convex, $\lim_{k \rightarrow \infty} \|\nabla \tilde{f}_l(x_k, \varepsilon_k)\| = 0$ implies $0 \in \partial f(\bar{x})$. $\square$

For the non-Lipschitz model ($l = 3$, $\varphi_3(t) = |t|^p$ with $0 < p < 1$), the regularizer is not locally Lipschitz at the origin. When difference components vanish ($d_i^T \bar{x} = 0$), the derivatives $\tilde{\mu}'_1(t, \varepsilon_k)$ and $\tilde{\mu}'_2(t, \varepsilon_k)$ can grow on the order of $O(\varepsilon_k^{p-1})$ for $|t| \leq \varepsilon_k$, so $\lim_{k \rightarrow \infty} \|\nabla \tilde{f}_3(x_k, \varepsilon_k)\| = 0$ alone does not guarantee limiting subdifferential stationarity ($0 \in \partial_L f(\bar{x})$) without additional componentwise scaling conditions [5]. Therefore, for $l = 3$, we adopt the weaker concept of *asymptotic ε -stationarity*: a point $\bar{x} \in \mathbb{R}^n$ is defined as an asymptotic ε -stationary point of (1) if there exist sequences $x_k \rightarrow \bar{x}$ and $\varepsilon_k \rightarrow 0^+$ such that

$$\lim_{k \rightarrow \infty} \|\nabla \tilde{f}_3(x_k, \varepsilon_k)\| = 0. \quad (22)$$

4 Algorithm and convergence results

In this section, we detail the Smoothing Conjugate Gradient Algorithm (SCGA) for solving the unconstrained smoothed problem (16) and establish its global convergence. To handle the dynamic reduction of the smoothing parameter without violating the descent property, SCGA is structured within an inner-outer continuation framework: during the j -th outer stage, $\varepsilon_j > 0$ remains fixed while the inner loop minimizes $\tilde{f}_l(x, \varepsilon_j)$ until the root-mean-square (RMS) gradient norm $\|g_k\|_{\text{rms}} := \|g_k\|/\sqrt{n}$ falls below τ_j (or an inner iteration cap $K_{\max}$ is reached), after which both ε_j and τ_j are systematically reduced. Denoting $g_k := \nabla \tilde{f}_l(x_k, \varepsilon_j)$ and $y_{k-1} := g_k - g_{k-1}$ at inner iteration k , the search direction d_k is constructed via the modified three-term Polak–Ribière–Polyak (PRP) scheme:

$$d_0 = -g_0, \quad d_k = -g_k + \beta_k d_{k-1} - \theta_k y_{k-1} \quad (k \geq 1), \quad \text{with} \quad \beta_k = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \theta_k = \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2}. \quad (23)$$

To prevent excessive direction growth, a standard safeguard resets the direction to the steepest descent vector ($d_k \leftarrow -g_k$) if $\|d_k\| > M_d \|g_k\|$ for a threshold $M_d > 1$. The step size $\alpha_k > 0$ is determined via an Armijo backtracking line search, as summarized in Algorithm 1.

Lemma 6. *For all inner iterations $k \geq 0$ of Algorithm 1, the search direction d_k defined in (23) satisfies the sufficient descent identity $g_k^T d_k = -\|g_k\|^2$. If the safeguard $d_k \leftarrow -g_k$ is triggered, $g_k^T d_k = -\|g_k\|^2$ holds trivially.*

Proof. For $k = 0$, $g_0^T d_0 = -\|g_0\|^2$. For $k \geq 1$, taking the inner product of g_k with d_k in (23) yields:

$$g_k^T d_k = -\|g_k\|^2 + \beta_k (g_k^T d_{k-1}) - \theta_k (g_k^T y_{k-1}) = -\|g_k\|^2 + \frac{(g_k^T y_{k-1})(g_k^T d_{k-1}) - (g_k^T d_{k-1})(g_k^T y_{k-1})}{\|g_{k-1}\|^2} = -\|g_k\|^2. \quad (25)$$

While this exact algebraic cancellation guarantees sufficient descent independently of line-search accuracy, the three-term structure also acts as an orthogonal projection onto the subspace orthogonal to g_k , which mitigates loss of conjugacy in ill-conditioned non-smooth problems. $\square$

Lemma 7. *For any fixed $\varepsilon_j > 0$, the Armijo line search (24) terminates in a finite number of steps m_k at every inner iteration $k \geq 0$, and the accepted step size satisfies:*

$$\alpha_k \geq \min \left\{ 1, c_j \frac{\|g_k\|^2}{\|d_k\|^2} \right\}, \quad \text{where } c_j := \frac{2(1-\sigma)\delta}{L_l(\varepsilon_j)} > 0. \quad (26)$$

Algorithm 1: Smoothing Conjugate Gradient Algorithm (SCGA) with Inner-Outer Continuation

Require: Starting point $x^{(0)} \in \mathbb{R}^n$, outer stage count $J \geq 1$, initial smoothing parameter $\varepsilon_0 > 0$, Armijo parameters $\sigma, \delta \in (0, 1)$, reduction factors $\tilde{\beta}, \gamma \in (0, 1)$, inner iteration cap $K_{\max} \in \mathbb{N} \cup \{\infty\}$, total iteration budget $N_{\max} \in \mathbb{N} \cup \{\infty\}$, and safeguard threshold $M_d > 1$.

Ensure: Restored image x^* and termination status flag $\text{status} \in \{\text{TOL}, \text{CAP}\}$.

- 1: Initialize outer stage $j \leftarrow 0$, total iteration counter $k_{\text{tot}} \leftarrow 0$, and $\text{status} \leftarrow \text{TOL}$.
- 2: **while** $j < J$ **and** $k_{\text{tot}} < N_{\max}$ **do**
- 3: Set $x_0 \leftarrow x^{(j)}$, $k \leftarrow 0$, compute $g_0 \leftarrow \nabla \tilde{f}_l(x_0, \varepsilon_j)$, and set $d_0 \leftarrow -g_0$.
- 4: **while** $\|g_k\|_{\text{rms}} \geq \tau_j$ **and** $k < K_{\max}$ **and** $k_{\text{tot}} < N_{\max}$ **do**
- 5: **if** $k \geq 1$ **then**
- 6: Compute $y_{k-1} = g_k - g_{k-1}$, $\beta_k = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}$, $\theta_k = \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2}$.
- 7: $d_k \leftarrow -g_k + \beta_k d_{k-1} - \theta_k y_{k-1}$.
- 8: **if** $\|d_k\| > M_d \|g_k\|$ **then**
- 9: $d_k \leftarrow -g_k$.
- 10: **end if**
- 11: **end if**
- 12: Find the smallest integer $m_k \geq 0$ such that $\alpha_k = \delta^{m_k}$ satisfies the Armijo condition:

$$\tilde{f}_l(x_k + \delta^{m_k} d_k, \varepsilon_j) \leq \tilde{f}_l(x_k, \varepsilon_j) + \sigma \delta^{m_k} g_k^T d_k. \quad (24)$$

- 13: Update $x_{k+1} \leftarrow x_k + \alpha_k d_k$, compute $g_{k+1} \leftarrow \nabla \tilde{f}_l(x_{k+1}, \varepsilon_j)$, and set $k \leftarrow k + 1$, $k_{\text{tot}} \leftarrow k_{\text{tot}} + 1$.
- 14: **end while**
- 15: Set $x^{(j+1)} \leftarrow x_k$, $\varepsilon_{j+1} \leftarrow \tilde{\beta} \varepsilon_j$, $\tau_{j+1} \leftarrow \gamma \tau_j$, and $j \leftarrow j + 1$.
- 16: **end while**
- 17: **if** $\|g_k\|_{\text{rms}} \geq \tau_{\text{final}}$ **then** set $\text{status} \leftarrow \text{CAP}$ **end if**
- 18: **return** $x^* = x^{(j)}$ and status .

Proof. By Lemma 4 and Lemma 6, the standard descent lemma gives $\tilde{f}_l(x_k + \alpha d_k, \varepsilon_j) \leq \tilde{f}_l(x_k, \varepsilon_j) + \alpha g_k^T d_k + \frac{L_l(\varepsilon_j)}{2} \alpha^2 \|d_k\|^2$. Thus, (24) holds whenever $\alpha \leq \frac{2(1-\sigma)\|g_k\|^2}{L_l(\varepsilon_j)\|d_k\|^2}$. Since $\delta \in (0, 1)$, backtracking terminates finitely. If $m_k = 0$, then $\alpha_k = 1$; if $m_k \geq 1$, the prior trial step $\alpha = \delta^{-1} \alpha_k$ violated (24), meaning $\delta^{-1} \alpha_k > \frac{2(1-\sigma)\|g_k\|^2}{L_l(\varepsilon_j)\|d_k\|^2}$, which immediately yields (26). $\square$

Theorem 4. Assume $\ker(A) = \{0\}$ and let Algorithm 1 be executed without artificial iteration caps ($K_{\max} = N_{\max} = \infty$) using the safeguard parameter $M_d > 1$. Then each inner loop terminates in finitely many steps. As $J \rightarrow \infty$ with $\tau_j \rightarrow 0$ and $\varepsilon_j \rightarrow 0$, the sequence $\{x^{(j)}\}$ satisfies $\lim_{j \rightarrow \infty} \|\nabla \tilde{f}_l(x^{(j)}, \varepsilon_j)\| = 0$. Furthermore, for the locally Lipschitz models ($l \in \{1, 2\}$), any accumulation point $\bar{x}$ of $\{x^{(j)}\}$ is a Clarke stationary point of the original problem (1). For the non-Lipschitz model ($l = 3$), $\bar{x}$ satisfies asymptotic ε -stationarity.

Proof. Fix an outer stage j with smoothing parameter $\varepsilon_j > 0$ and tolerance $\tilde{\tau}_j := \sqrt{n} \tau_j > 0$. Suppose, for the sake of contradiction, that the inner loop does not terminate finitely, meaning that $\|g_k\| \geq \tilde{\tau}_j > 0$

holds for all $k \geq 0$. Combining the Armijo condition (24) with the sufficient descent identity (??) yields:

$$\tilde{f}_l(x_k, \varepsilon_j) - \tilde{f}_l(x_{k+1}, \varepsilon_j) \geq -\sigma \alpha_k g_k^T d_k = \sigma \alpha_k \|g_k\|^2 > 0, \quad \forall k \geq 0. \quad (27)$$

Consequently, the sequence $\{\tilde{f}_l(x_k, \varepsilon_j)\}$ is strictly decreasing, and $\{x_k\} \subset \mathcal{L}(x_0, \varepsilon_j)$. By Lemma 5, $\mathcal{L}(x_0, \varepsilon_j)$ is compact, so $\tilde{f}_l(\cdot, \varepsilon_j)$ is bounded below on $\mathcal{L}(x_0, \varepsilon_j)$ by a finite constant $\tilde{f}_{\min}$. Summing (27) over $k \geq 0$ yields:

$$\sigma \sum_{k=0}^{\infty} \alpha_k \|g_k\|^2 \leq \tilde{f}_l(x_0, \varepsilon_j) - \tilde{f}_{\min} < \infty \implies \sum_{k=0}^{\infty} \alpha_k \|g_k\|^2 < \infty. \quad (28)$$

Substituting the step-size lower bound (26) from Lemma 7 into (28) gives:

$$\sum_{k=0}^{\infty} \min \left\{ \|g_k\|^2, c_j \frac{\|g_k\|^4}{\|d_k\|^2} \right\} \leq \sum_{k=0}^{\infty} \alpha_k \|g_k\|^2 < \infty. \quad (29)$$

Since $\|g_k\|^2 \geq \tilde{\tau}_j^2 > 0$, the first term in the minimum cannot vanish. Thus, for sufficiently large k , the minimum is attained by the second term, establishing the Zoutendijk condition:

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty, \quad \text{which implies} \quad \tilde{\tau}_j^4 \sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} \leq \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \implies \sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} < \infty. \quad (30)$$

Next, we bound the search direction norm $\|d_k\|$. Since $\nabla \tilde{f}_l(\cdot, \varepsilon_j)$ is continuous on the compact level set $\mathcal{L}(x_0, \varepsilon_j)$, there exists $\Gamma_j > 0$ such that $\|g_k\| \leq \Gamma_j$ for all $k \geq 0$. By the explicit safeguard enforced in Algorithm 1, the search direction strictly satisfies $\|d_k\| \leq M_d \|g_k\|$ (either natively or via the reset $d_k = -g_k$). Therefore, we obtain the uniform upper bound:

$$\|d_k\| \leq M_d \|g_k\| \leq M_d \Gamma_j =: D_j < \infty, \quad \forall k \geq 0. \quad (31)$$

However, (31) implies $\sum_{k=0}^{\infty} \|d_k\|^{-2} \geq \sum_{k=0}^{\infty} D_j^{-2} = \infty$, which directly contradicts the reciprocal summability derived in (30). Thus, the inner loop terminates in a finite number of iterations with $\|g_k\|_{\text{rms}} < \tau_j$.

In the outer loop, $x^{(j+1)} = x_k$ satisfies $\|\nabla \tilde{f}_l(x^{(j+1)}, \varepsilon_j)\| < \sqrt{n} \tau_j$. As $j \rightarrow \infty$, both $\tau_j \rightarrow 0$ and $\varepsilon_j \rightarrow 0$, so $\lim_{j \rightarrow \infty} \|\nabla \tilde{f}_l(x^{(j+1)}, \varepsilon_j)\| = 0$. For the locally Lipschitz models ($l \in \{1, 2\}$), Theorem 3 guarantees that any accumulation point $\bar{x}$ satisfies $0 \in \partial f(\bar{x})$. For the non-Lipschitz model ($l = 3$), the condition ensures $\bar{x}$ satisfies asymptotic ε -stationarity. $\square$

Remark 1. While Theorem 4 assumes asymptotic convergence ($K_{\max} = J = \infty$), solving early continuation stages to high precision is computationally redundant and risks overfitting. Thus, our implementation employs a finite three-stage schedule ($J = 3$) with an inner iteration cap ($K_{\max}$) and records a termination status flag (TOL or CAP) to bridge theory and practice.

5 Numerical experiments

In this section, we evaluate the computational efficiency and restoration accuracy of the proposed SCGA framework on standard image restoration problems in MATLAB R2025b (Intel Core i7, 8 GB RAM). To ensure strict reproducibility and reliability, each degradation scenario is executed over $M = 5$ independent noise realizations using fixed seeds $\text{rng}(1000 \cdot i_{\text{img}} + 100 \cdot i_{\text{deg}} + s)$ with $s \in \{101, 202, 303, 404, 505\}$,

and results are reported as Mean $\pm$ Std. Three benchmark images (*Cameraman*, *Peppers*, and *Coins*) are standardized to 256×256 pixels with intensities in $[0, 1]$ ($n = 65536$). The discrete difference matrix $D \in \mathbb{R}^{r \times n}$ ($r = 2n = 131072$) stacks the first-order horizontal and vertical forward finite-difference operators under periodic boundary conditions, with D^T denoting the negative discrete divergence. Degraded observations follow $b = Ax + n$, where $A = (1 - \gamma_{\text{reg}})H_{\text{blur}} + \gamma_{\text{reg}}I$ ($\gamma_{\text{reg}} = 10^{-2}$) regularizes a 5×5 Gaussian kernel (standard deviation 0.75, periodic boundaries) to satisfy $\ker(A) = \{0\}$ in Lemma 5. Three noise regimes are tested: moderate Gaussian noise ($\sigma = 0.05$), severe Gaussian noise ($\sigma = 0.10$), and Salt & Pepper impulse noise ($d = 0.05$). Since the quadratic fidelity term in (1) is tailored for Gaussian noise, impulse noise is handled via a standard two-phase procedure where outlier pixels ($b_i \in \{0, 1\}$) are pre-filtered by a 3×3 symmetric median filter prior to optimization.

All methods are initialized at $x^{(0)} = b$ (or the pre-filtered image for impulse noise). The outer continuation loop runs over $J = 3$ stages with $\varepsilon_j \in \{0.04, 0.015, 0.005\}$ and inner RMS gradient tolerances $\tau_j \in \{2.5 \times 10^{-2}, 1.5 \times 10^{-2}, 1.2 \times 10^{-2}\}$, subject to a per-stage inner cap of $K_{\max} = 8$ (with at least 4 inner steps per stage) and a total budget of $N_{\max} = 100$. In Table 1, the **Status** column explicitly records **TOL** if the final target tolerance $\|g_k\|_{\text{rms}} = \|\nabla \tilde{f}_l(x_k, \varepsilon_{\text{final}})\|/\sqrt{n} < \tau_{\text{final}} = 1.2 \times 10^{-2}$ is satisfied, or **CAP** if $K_{\max}$ is reached first (Remark 1). While τ_{final} controls numerical stationarity, the theoretical smoothing discrepancy $|f(x) - \tilde{f}_l(x, \varepsilon_{\text{final}})|$ is bounded by $\varepsilon_{\text{final}} = 0.005$ via Theorem 1. We set $\sigma = 10^{-4}$ and $\delta = 0.5$ in (24), $\lambda = 0.008 + 0.14\sigma$ for Gaussian noise ($\lambda = 0.011$ for impulse noise), $\alpha = 0.45$ in φ_2 , and $p = 0.85$ in φ_3 . Performance is measured via mean iterations (**Iter**), CPU time (**CPU (s)**), function/gradient/line-search evaluations (**f_eval**, **g_eval**, **ls_eval**), final RMS gradient norm ($\|g_k\|_{\text{rms}}$), **Status**, **SSIM**, and **PSNR (dB)**:

$$\text{PSNR}(x_{\text{true}}, x^*) = 10 \log_{10} \left(\frac{n}{\|x_{\text{true}} - x^*\|^2} \right), \quad (\text{MAX}_I = 1). \quad (32)$$

To separate the effect of the three-term SCGA solver from the smoothing and continuation choices, we evaluate nine methods: five proposed variants—**SCGA with TV-LS** ($\varphi_1, \tilde{\rho}_1^{(1)}$), **TV-GS** ($\varphi_1, \tilde{\rho}_2$), **NP-LS** ($\varphi_2, \tilde{\rho}_1^{(1)}$), **NP-GS** ($\varphi_2, \tilde{\rho}_2$), and **NLP-S** ($\varphi_3, \tilde{\mu}_1, p = 0.85$); two matched-continuation baselines—**PRP-Cont (TV-LS)** [31, 32] and **FISTA-Cont (TV-LS)** [3], which minimize the exact same locally smoothed TV objective ($\tilde{\rho}_1^{(1)}$) under the identical ε_j and τ_j schedule; and two fixed-smoothing baselines—**PRP (Fixed)** and **FISTA (Fixed)**, which use classical hyperbolic smoothing $\sqrt{t^2 + \varepsilon^2}$ with fixed $\varepsilon = 0.08$.

To isolate the algorithmic contribution of the three-term SCGA solver from the effects of the smoothing functions and the continuation schedule, five proposed SCGA variants are compared against both matched-continuation baselines and fixed-smoothing baselines:

- **SCGA with TV-LS / TV-GS**: Total Variation model (φ_1) with local polynomial smoothing $\tilde{\rho}_1^{(1)}(t, \varepsilon)$ and global arctangent smoothing $\tilde{\rho}_2(t, \varepsilon)$, respectively.
- **SCGA with NP-LS / NP-GS**: Nonconvex Lipschitz model (φ_2) with local smoothing $\tilde{\rho}_1^{(1)}(t, \varepsilon)$ and global smoothing $\tilde{\rho}_2(t, \varepsilon)$, respectively.
- **SCGA with NLP-S**: Nonconvex Non-Lipschitz model ($\varphi_3, p = 0.85$) with C^1 local smoothing $\tilde{\mu}_1(t, \varepsilon)$.
- **PRP-Cont (TV-LS)** and **FISTA-Cont (TV-LS)** (Matched-Continuation Baselines): Standard two-term Polak–Ribière–Polyak CG [31, 32] and Fast Iterative Shrinkage-Thresholding Algorithm [3]



Figure 1: Visual comparison of degraded and restored *Cameraman* images under Gaussian ($\sigma = 0.05$), Gaussian ($\sigma = 0.10$), and Salt & Pepper ($d = 0.05$) noise

applied to the exact same locally smoothed TV objective ($\tilde{\rho}_1^{(1)}$) using the identical three-stage continuation schedule $\epsilon_j \in \{0.04, 0.015, 0.005\}$ and tolerances τ_j .

- **PRP (Fixed)** and **FISTA (Fixed)** (Fixed-Smoothing Baselines): Standard two-term PRP and FISTA utilizing classical hyperbolic smoothing $\tilde{\varphi}(t, \epsilon) = \sqrt{t^2 + \epsilon^2}$ with a fixed smoothing parameter $\epsilon = 0.08$.

The results in Table 1 and Figures 1–3 indicate the following:

Table 1: Quantitative comparison (Mean $\pm$ Std over 5 noise realizations) of the proposed SCGA variants against matched-continuation and fixed-smoothing baselines

Image	Degradation	Method	Noisy PSNR	Iter	CPU (s)	f_eval	g_eval	ls_eval	$\ g_k\ _{rms}$	Status	PSNR (dB)	SSIM	
Camel	Gaussian ($\sigma = 0.05$)	SCGA with TV-LS	14.0	0.2617	35.0	17.0	18.0	1.16e-02	TOL	27.434 $\pm$ 0.033	0.8143 $\pm$ 0.0010		
		SCGA with TV-GS	17.0	0.2664	43.8	20.0	23.8	1.28e-02	CAP	27.497 $\pm$ 0.034	0.8117 $\pm$ 0.0015		
		SCGA with NP-LS	18.0	0.2979	49.4	21.0	28.4	1.92e-02	CAP	27.431 $\pm$ 0.037	0.8203 $\pm$ 0.0013		
		SCGA with NP-GS	18.2	0.3261	50.2	21.2	29.0	1.10e-02	TOL	27.506 $\pm$ 0.040	0.8196 $\pm$ 0.0007		
		SCGA with NLP-S	12.0	0.4560	28.0	15.0	13.0	4.63e-03	TOL	27.398 $\pm$ 0.042	0.8045 $\pm$ 0.0011		
		PRP-Cont (TV-LS)	18.2	0.3854	82.0	21.2	60.8	1.01e-02	TOL	27.421 $\pm$ 0.034	0.8160 $\pm$ 0.0014		
		FISTA-Cont (TV-LS)	21.0	0.1421	23.0	23.0	0.0	1.32e-02	CAP	26.703 $\pm$ 0.029	0.7230 $\pm$ 0.0008		
		PRP (Fixed)	20.0	0.2128	39.0	20.0	19.0	2.50e-04	TOL	26.875 $\pm$ 0.023	0.7279 $\pm$ 0.0009		
		FISTA (Fixed)	24.0	0.1107	24.0	24.0	0.0	6.30e-04	TOL	26.882 $\pm$ 0.023	0.7275 $\pm$ 0.0009		
		SCGA with TV-LS	20.6	0.3317	52.4	23.6	28.8	1.81e-02	CAP	25.017 $\pm$ 0.058	0.6626 $\pm$ 0.0058		
		SCGA with TV-GS	24.0	0.3099	56.4	27.0	29.4	1.63e-02	CAP	24.870 $\pm$ 0.058	0.6428 $\pm$ 0.0054		
		SCGA with NP-LS	24.0	0.3628	61.2	27.0	34.2	2.01e-02	CAP	25.059 $\pm$ 0.057	0.6813 $\pm$ 0.0090		
	Gaussian ($\sigma = 0.10$)	SCGA with NP-GS	24.0	0.3729	59.8	27.0	32.8	2.35e-02	CAP	24.963 $\pm$ 0.055	0.6686 $\pm$ 0.0058		
		SCGA with NLP-S	14.0	0.4804	32.0	17.0	15.0	1.01e-02	TOL	25.077 $\pm$ 0.048	0.6793 $\pm$ 0.0048		
		PRP-Cont (TV-LS)	24.0	0.5137	115.8	27.0	88.8	2.44e-02	CAP	24.857 $\pm$ 0.063	0.6389 $\pm$ 0.0059		
		FISTA-Cont (TV-LS)	24.0	0.1368	24.0	24.0	0.0	2.47e-02	CAP	22.808 $\pm$ 0.037	0.4361 $\pm$ 0.0024		
		PRP (Fixed)	20.0	0.1979	39.0	20.0	19.0	2.10e-04	TOL	24.540 $\pm$ 0.048	0.5667 $\pm$ 0.0032		
		FISTA (Fixed)	24.0	0.1172	24.0	24.0	0.0	2.50e-03	TOL	24.625 $\pm$ 0.048	0.5767 $\pm$ 0.0031		
	Salt & Pepper ($d = 0.05$)	SCGA with TV-LS	12.2	0.1752	31.0	15.2	15.8	9.03e-03	TOL	28.535 $\pm$ 0.016	0.8594 $\pm$ 0.0008		
		SCGA with TV-GS	14.6	0.2751	36.8	17.6	19.2	1.03e-02	TOL	28.815 $\pm$ 0.020	0.8660 $\pm$ 0.0002		
		SCGA with NP-LS	13.6	0.2608	35.2	16.6	18.6	1.01e-02	TOL	28.544 $\pm$ 0.018	0.8563 $\pm$ 0.0004		
		SCGA with NP-GS	14.6	0.3086	39.2	17.6	21.6	1.15e-02	TOL	28.669 $\pm$ 0.055	0.8608 $\pm$ 0.0013		
		SCGA with NLP-S	12.0	0.4801	28.0	15.0	13.0	2.70e-03	TOL	28.763 $\pm$ 0.022	0.8713 $\pm$ 0.0002		
		PRP-Cont (TV-LS)	16.0	0.3252	68.6	19.0	49.6	8.96e-03	TOL	28.518 $\pm$ 0.014	0.8599 $\pm$ 0.0004		
FISTA-Cont (TV-LS)		18.0	0.1225	21.0	21.0	0.0	6.80e-03	TOL	28.223 $\pm$ 0.015	0.8587 $\pm$ 0.0004			
PRP (Fixed)		20.0	0.1708	39.0	20.0	19.0	3.00e-04	TOL	28.653 $\pm$ 0.020	0.8890 $\pm$ 0.0001			
FISTA (Fixed)		24.0	0.1222	24.0	24.0	0.0	3.60e-04	TOL	28.715 $\pm$ 0.022	0.8899 $\pm$ 0.0001			
SCGA with TV-LS		15.0	0.1984	37.0	18.0	19.0	6.67e-03	TOL	27.407 $\pm$ 0.031	0.8361 $\pm$ 0.0014			
SCGA with TV-GS		16.0	0.2313	40.0	19.0	21.0	1.11e-02	TOL	27.486 $\pm$ 0.046	0.8333 $\pm$ 0.0046			
SCGA with NP-LS		20.8	0.3407	53.2	23.8	29.4	9.44e-03	TOL	27.520 $\pm$ 0.032	0.8393 $\pm$ 0.0017			
Gaussian ($\sigma = 0.05$)	SCGA with NP-GS	19.0	0.3430	48.0	22.0	26.0	1.07e-02	TOL	27.543 $\pm$ 0.039	0.8380 $\pm$ 0.0012			
	SCGA with NLP-S	12.0	0.4546	28.0	15.0	13.0	4.59e-03	TOL	27.515 $\pm$ 0.033	0.8392 $\pm$ 0.0012			
	PRP-Cont (TV-LS)	17.0	0.4077	80.4	20.0	60.4	1.06e-02	TOL	27.357 $\pm$ 0.031	0.8324 $\pm$ 0.0015			
	FISTA-Cont (TV-LS)	20.8	0.1345	22.8	22.8	0.0	1.29e-02	CAP	26.612 $\pm$ 0.035	0.7566 $\pm$ 0.0032			
	PRP (Fixed)	20.0	0.1627	39.0	20.0	19.0	1.60e-04	TOL	26.899 $\pm$ 0.024	0.7786 $\pm$ 0.0018			
	FISTA (Fixed)	24.0	0.1075	24.0	24.0	0.0	6.10e-04	TOL	26.904 $\pm$ 0.024	0.7784 $\pm$ 0.0018			
	Peppers	Gaussian ($\sigma = 0.10$)	SCGA with TV-LS	23.8	0.2778	56.0	26.8	29.2	1.25e-02	CAP	24.868 $\pm$ 0.093	0.6933 $\pm$ 0.0033	
			SCGA with TV-GS	23.6	0.3132	57.2	26.6	30.6	9.14e-03	TOL	24.696 $\pm$ 0.073	0.6711 $\pm$ 0.0026	
			SCGA with NP-LS	23.8	0.3854	62.4	26.8	35.6	1.59e-02	CAP	24.917 $\pm$ 0.080	0.7050 $\pm$ 0.0022	
			SCGA with NP-GS	24.0	0.3699	59.8	27.0	32.8	2.13e-02	CAP	24.826 $\pm$ 0.083	0.6937 $\pm$ 0.0022	
			SCGA with NLP-S	15.0	0.5600	36.0	18.0	18.0	4.70e-03	TOL	25.028 $\pm$ 0.083	0.7177 $\pm$ 0.0022	
			PRP-Cont (TV-LS)	24.0	0.4832	110.4	27.0	83.4	2.11e-02	CAP	24.825 $\pm$ 0.091	0.6822 $\pm$ 0.0051	
Salt & Pepper ($d = 0.05$)		FISTA-Cont (TV-LS)	24.0	0.1408	24.0	24.0	0.0	2.51e-02	CAP	22.621 $\pm$ 0.025	0.4796 $\pm$ 0.0011		
		PRP (Fixed)	20.0	0.1825	39.0	20.0	19.0	1.60e-04	TOL	24.545 $\pm$ 0.049	0.6288 $\pm$ 0.0019		
		FISTA (Fixed)	24.0	0.1178	24.0	24.0	0.0	2.57e-03	TOL	24.655 $\pm$ 0.053	0.6401 $\pm$ 0.0019		
		SCGA with TV-LS	12.4	0.1479	30.2	15.4	14.8	7.57e-03	TOL	28.894 $\pm$ 0.053	0.8936 $\pm$ 0.0005		
		SCGA with TV-GS	12.0	0.1620	29.0	15.0	14.0	9.06e-03	TOL	29.062 $\pm$ 0.049	0.8989 $\pm$ 0.0003		
		SCGA with NP-LS	12.0	0.1950	31.0	15.0	16.0	7.84e-03	TOL	28.841 $\pm$ 0.044	0.8917 $\pm$ 0.0003		
Salt & Pepper ($d = 0.05$)	SCGA with NP-GS	13.0	0.2202	35.0	16.0	19.0	1.06e-02	TOL	29.000 $\pm$ 0.042	0.8951 $\pm$ 0.0003			
	SCGA with NLP-S	12.0	0.4216	27.0	15.0	12.0	2.36e-03	TOL	29.245 $\pm$ 0.060	0.9029 $\pm$ 0.0009			
	PRP-Cont (TV-LS)	15.0	0.2918	64.0	18.0	46.0	9.58e-03	TOL	28.877 $\pm$ 0.042	0.8936 $\pm$ 0.0003			
	FISTA-Cont (TV-LS)	18.0	0.1131	21.0	21.0	0.0	6.17e-03	TOL	28.469 $\pm$ 0.039	0.8924 $\pm$ 0.0003			
	PRP (Fixed)	20.0	0.1597	39.0	20.0	19.0	2.10e-04	TOL	28.857 $\pm$ 0.033	0.9111 $\pm$ 0.0002			
	FISTA (Fixed)	24.0	0.1057	24.0	24.0	0.0	3.10e-04	TOL	28.950 $\pm$ 0.034	0.9119 $\pm$ 0.0002			
	Gaussian ($\sigma = 0.05$)	SCGA with TV-LS	14.0	0.1775	35.0	17.0	18.0	1.13e-02	TOL	31.182 $\pm$ 0.054	0.8799 $\pm$ 0.0022		
		SCGA with TV-GS	17.0	0.3039	44.0	20.0	24.0	1.31e-02	CAP	31.113 $\pm$ 0.059	0.8736 $\pm$ 0.0023		
		SCGA with NP-LS	18.0	0.2872	49.6	21.0	28.6	1.69e-02	CAP	31.114 $\pm$ 0.064	0.8893 $\pm$ 0.0015		
		SCGA with NP-GS	18.4	0.3200	49.8	21.4	28.4	1.09e-02	TOL	31.107 $\pm$ 0.061	0.8896 $\pm$ 0.0034		
		SCGA with NLP-S	12.0	0.4283	28.0	15.0	13.0	3.91e-03	TOL	30.983 $\pm$ 0.056	0.8670 $\pm$ 0.0019		
		PRP-Cont (TV-LS)	17.0	0.3508	78.0	20.0	58.0	8.92e-03	TOL	31.117 $\pm$ 0.055	0.8758 $\pm$ 0.0022		
FISTA-Cont (TV-LS)		21.0	0.1910	23.0	23.0	0.0	1.35e-02	CAP	29.626 $\pm$ 0.034	0.7477 $\pm$ 0.0022			
PRP (Fixed)		20.0	0.1567	39.0	20.0	19.0	1.50e-04	TOL	30.173 $\pm$ 0.028	0.7627 $\pm$ 0.0017			
FISTA (Fixed)		24.0	0.1025	24.0	24.0	0.0	5.90e-04	TOL	30.184 $\pm$ 0.027	0.7621 $\pm$ 0.0017			
Coins		Gaussian ($\sigma = 0.10$)	SCGA with TV-LS	24.0	0.3049	57.2	27.0	30.2	2.66e-02	CAP	27.386 $\pm$ 0.053	0.6965 $\pm$ 0.0024	
			SCGA with TV-GS	23.6	0.3345	56.0	26.6	29.4	1.00e-02	TOL	26.939 $\pm$ 0.061	0.6586 $\pm$ 0.0028	
			SCGA with NP-LS	22.8	0.4056	60.4	25.8	34.6	2.05e-02	CAP	27.384 $\pm$ 0.063	0.7141 $\pm$ 0.0050	
	SCGA with NP-GS		24.0	0.4389	61.0	27.0	34.0	2.19e-02	CAP	27.162 $\pm$ 0.061	0.6941 $\pm$ 0.0026		
	SCGA with NLP-S		16.0	0.6095	37.2	19.0	18.2	4.66e-03	TOL	27.537 $\pm$ 0.063	0.7265 $\pm$ 0.0025		
	PRP-Cont (TV-LS)		23.8	0.5283	112.4	26.8	85.6	2.03e-02	CAP	27.092 $\pm$ 0.124	0.6648 $\pm$ 0.0116		
	Salt & Pepper ($d = 0.05$)	FISTA-Cont (TV-LS)	24.0	0.1397	24.0	24.0	0.0	2.59e-02	CAP	23.668 $\pm$ 0.039	0.4018 $\pm$ 0.0018		
		PRP (Fixed)	20.0	0.1670	39.0	20.0	19.0	1.30e-04	TOL	26.592 $\pm$ 0.038	0.5747 $\pm$ 0.0016		
		FISTA (Fixed)	24.0	0.1162	24.0	24.0	0.0	2.58e-03	TOL	26.792 $\pm$ 0.039	0.5892 $\pm$ 0.0017		
		SCGA with TV-LS	12.0	0.1549	31.0	15.0	16.0	6.04e-03	TOL	33.758 $\pm$ 0.039	0.9408 $\pm$ 0.0004		
		SCGA with TV-GS	13.0	0.1842	33.6	16.0	17.6	1.09e-02	TOL	34.135 $\pm$ 0.038	0.9456 $\pm$ 0.0004		
		SCGA with NP-LS	12.0	0.2150	33.0	15.0	18.0	5.48e-03	TOL	33.641 $\pm$ 0.030	0.9383 $\pm$ 0.0003		
Salt & Pepper ($d = 0.05$)	SCGA with NP-GS	16.0	0.2619	41.6	19.0	22.6	1.12e-02	TOL	33.830 $\pm$ 0.046	0.9417 $\pm$ 0.0003			
	SCGA with NLP-S	12.0	0.4140	27.0	15.0	12.0	2.04e-03	TOL	33.731 $\pm$ 0.024	0.9425 $\pm$ 0.0002			
	PRP-Cont (TV-LS)	15.0	0.2878	63.4	18.0	45.4	7.87e-03	TOL	33.740 $\pm$ 0.040	0.9408 $\pm$ 0.0004			
	FISTA-Cont (TV-LS)	18.0	0.1131	21.0	21.0	0.0	7.47e-03	TOL	33.449 $\pm$ 0.040	0.9401 $\pm$ 0.0003			
	PRP (Fixed)	20.0	0.1528	39.0	20.0	19.0	1.90e-04	TOL	34.360 $\pm$ 0.045	0.9578 $\pm$ 0.0003			
	FISTA (Fixed)	24.0	0.1048	24.0	24.0	0.0	3.00e-04	TOL	34.422 $\pm$ 0.044	0.9580 $\pm$ 0.0003			



Figure 2: Visual comparison of degraded and restored *Peppers* images under Gaussian ($\sigma = 0.05$), Gaussian ($\sigma = 0.10$), and Salt & Pepper ($d = 0.05$) noise

the continuation scheme increases Gaussian restoration PSNR and SSIM. Second, the performance of the regularizers varies with the noise level. Under moderate Gaussian noise ($\sigma = 0.05$), the non-convex Lipschitz models (**NP-GS** and **NP-LS**) yield higher PSNR on *Cameraman* and *Peppers*, and higher SSIM on all three images. **TV-LS** produces the highest PSNR on *Coins*. Under severe Gaussian noise ($\sigma = 0.10$), the non-Lipschitz model (**NLP-S**) attains the highest PSNR across all images and the highest SSIM on *Peppers* and *Coins*. Finally, the stopping status and metrics reveal practical trade-offs. At $\sigma = 0.05$ and $d = 0.05$, the SCGA variants generally satisfy the tolerance (**TOL**) within 12.0–20.8 iterations. At $\sigma = 0.10$, Lipschitz variants reach the stage cap (**CAP**), whereas **NLP-S** satisfies **TOL** in

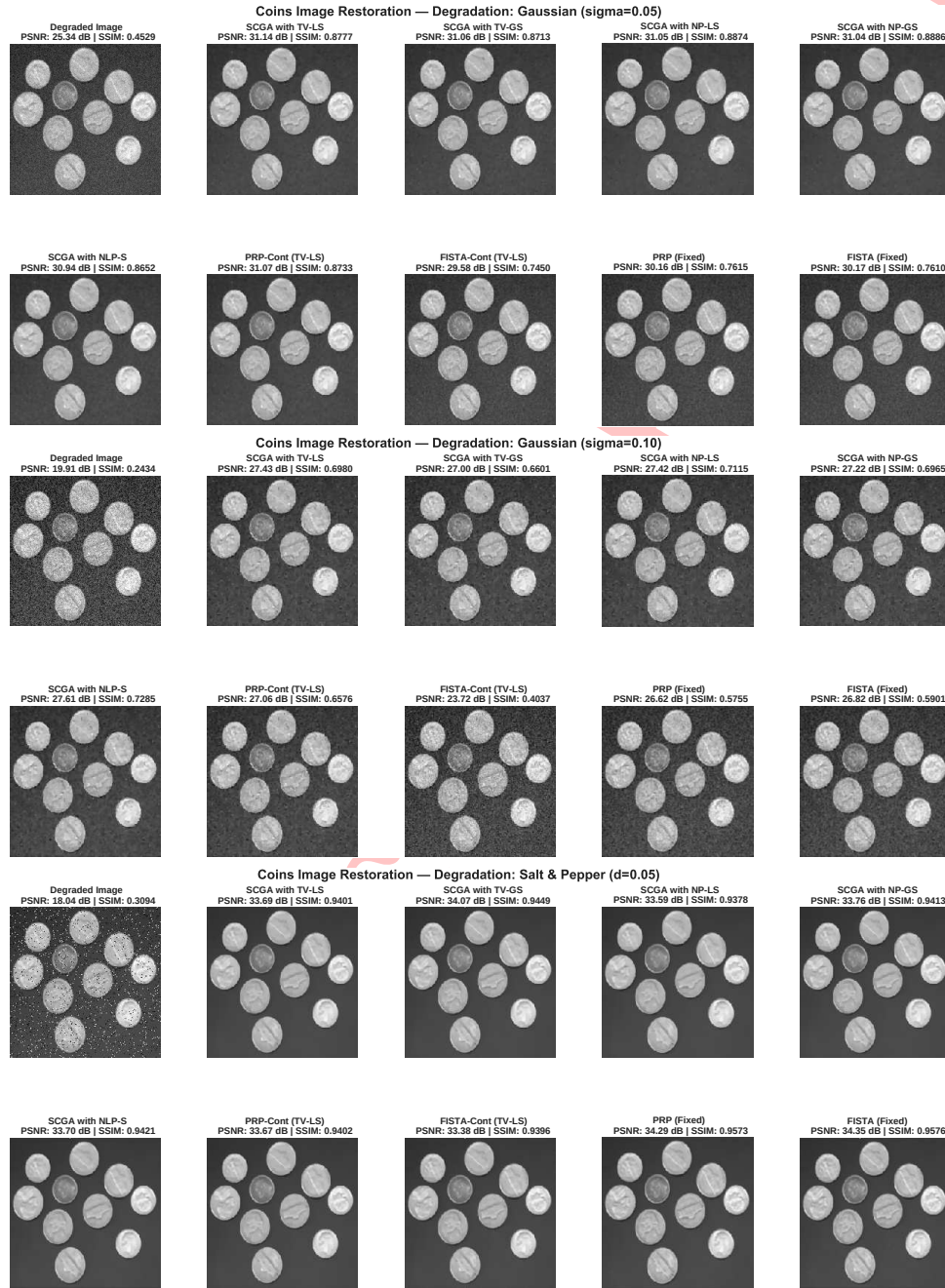


Figure 3: Visual comparison of degraded and restored *Coins* images under Gaussian ($\sigma = 0.05$), Gaussian ($\sigma = 0.10$), and Salt & Pepper ($d = 0.05$) noise

14.0–16.0 iterations. In the impulse noise tests, **SCGA with TV-GS** and **NLP-S** yield higher PSNR on *Cameraman* and *Peppers*. Conversely, **FISTA (Fixed)** and **PRP (Fixed)** produce higher SSIM across all images, higher PSNR on the piecewise-smooth *Coins* image, and lower CPU times. For **NLP-S**,

the fractional power evaluations result in fewer function evaluations but higher CPU times. Standard deviations across the five realizations remain small.

6 Conclusion

In this paper, we integrated indicator-based smoothing functions (local polynomial and global arctangent) with a three-term Polak–Ribière–Polyak conjugate gradient algorithm (SCGA) using an inner-outer continuation scheme. This approach reformulated Total Variation, non-convex Lipschitz, and non-Lipschitz regularizers into continuously differentiable optimization problems with explicit $O(\varepsilon)$ and $O(\varepsilon^p)$ uniform approximation error bounds. The three-term search direction satisfies the sufficient descent condition unconditionally. By holding the smoothing parameter constant during the Armijo line search, applying a direction safeguard, and reducing ε_j across stages, we established finite inner-loop termination. Under the injectivity condition $\ker(A) = \{0\}$, accumulation points are Clarke stationary for the locally Lipschitz models and satisfy asymptotic ε -stationarity for the non-Lipschitz model. Numerical experiments over five independent noise realizations demonstrated practical trade-offs. Under matched-continuation schedules, SCGA required fewer line-search evaluations and lower CPU time compared to PRP-Cont (TV-LS). The non-convex Lipschitz variants yielded higher PSNR under moderate Gaussian noise ($\sigma = 0.05$), whereas the non-Lipschitz variant (NLP-S) produced higher PSNR under severe noise ($\sigma = 0.10$). In impulse noise settings, fixed-smoothing baselines maintained higher SSIM and lower computational times. Future work will extend the coercivity analysis to rank-deficient kernels via $\ker(A) \cap \ker(D) = \{0\}$. Additionally, replacing the quadratic fidelity term with a smoothed ℓ_1 norm to handle impulse noise in a single stage, along with automated parameter selection for λ , α , and p , remain topics for future investigation.

Data Availability and Acknowledgements

The complete MATLAB benchmark scripts, random seeds, and data-generation settings are available from the corresponding author upon request and are provided in the supplementary material. During the preparation of this work, the authors used generative AI tools solely to improve language, grammar, and overall readability.

Conflicts of interest

The author declares that there are no conflicts of interest.

References

- [1] M.V. Afonso, J.M. Bioucas-Dias, M.A.T. Figueiredo, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process. **19** (2010) 2345–2356.
- [2] A.M. Bagirov, A. Al Nuaimat, N. Sultanova, *Hyperbolic smoothing function method for minimax problems*, Optimization **62** (2013) 759–782.
- [3] A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. **2** (2009) 183–202.
- [4] S. Becker, J. Bobin, E.J. Candès, *NESTA: A fast and accurate first-order method for sparse recovery*, SIAM J. Imaging Sci. **4** (2011) 1–39.

- [5] W. Bian, X. Chen, *Linearly constrained non-Lipschitz optimization for image restoration*, SIAM J. Imaging Sci. **8** (2015) 2294–2322.
- [6] J. Cao, J. Wu, *A conjugate gradient algorithm and its applications in image restoration*, Appl. Numer. Math. **152** (2020) 243–252.
- [7] C. Chen, O.L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comput. Optim. Appl. **5** (1996) 97–138.
- [8] X. Chen, W. Zhou, *Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization*, SIAM J. Imaging Sci. **3** (2010) 765–790.
- [9] X. Chen, M.K. Ng, C. Zhang, *Non-Lipschitz ℓ_p -regularization and box constrained model for image restoration*, IEEE Trans. Image Process. **21** (2012) 4709–4721.
- [10] X. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program. **134** (2012) 71–99.
- [11] Y.H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim. **10** (1999) 177–182.
- [12] I. Daubechies, M. Defrise, C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pure Appl. Math. **57** (2004) 1413–1457.
- [13] R. Fletcher, C.M. Reeves, *Function minimization by conjugate gradients*, Comput. J. **7** (1964) 149–154.
- [14] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, 2000.
- [15] G.H. Golub, P.C. Hansen, D.P. O’Leary, *Tikhonov regularization and total least squares*, SIAM J. Matrix Anal. Appl. **21** (1999) 185–194.
- [16] E.T. Hale, W. Yin, Y. Zhang, *Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence*, SIAM J. Optim. **19** (2008) 1107–1130.
- [17] M.R. Hestenes, E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand. **49** (1952) 409–436.
- [18] Y. Huang, H. Liu, W. Cong, *A note on the smoothing quadratic regularization method for non-Lipschitz optimization*, Numer. Algorithms **69** (2015) 863–874.
- [19] X. Jiang, X. Ye, Z. Huang, M. Liu, *A family of hybrid conjugate gradient method with restart procedure for unconstrained optimizations and image restorations*, Comput. Oper. Res. **159** (2023) 106341.
- [20] X. Jiang, H. Yang, J. Yin, W. Liao, *A three-term conjugate gradient algorithm with restart procedure to solve image restoration problems*, J. Comput. Appl. Math. **424** (2023) 115020.
- [21] G. Landi, *A modified Newton projection method for ℓ_1 -regularized least squares image deblurring*, J. Math. Imaging Vis. **51** (2015) 195–208.

- [22] Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms*, J. Optim. Theory Appl. **69** (1991) 129–137.
- [23] N. Mau Nam, L.T. Hoai An, D. Giles, N. Thai An, *Smoothing techniques and difference of convex functions algorithms for image reconstructions*, Optimization **69** (2020) 1601–1633.
- [24] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program. **103** (2005) 127–152.
- [25] M.K. Ng, R.H. Chan, W.C. Tang, *A fast algorithm for deblurring models with Neumann boundary conditions*, SIAM J. Sci. Comput. **21** (1999) 851–866.
- [26] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal. **40** (2002) 965–994.
- [27] M. Nikolova, M.K. Ng, S. Zhang, W.K. Ching, *Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization*, SIAM J. Imaging Sci. **1** (2008) 2–25.
- [28] C.T. Pham, G. Gamard, A. Kopylov, T.T.T. Tran, *An algorithm for image restoration with mixed noise using total variation regularization*, Turk. J. Elec. Eng. Comput. Sci. **26** (2018) 2831–2845.
- [29] C.T. Pham, T.T.T. Tran, G. Gamard, *An efficient total variation minimization method for image restoration*, Informatica **31** (2020) 539–560.
- [30] C.T. Pham, T.T.T. Tran, H.V. Dang, H.P. Dang, *An adaptive image restoration algorithm based on hybrid total variation regularization*, Turk. J. Elec. Eng. Comput. Sci. **31** (2023) 1–16.
- [31] E. Polak, G. Ribière, *Note sur la convergence de méthodes de directions conjuguées*, Rev. Fr. Inform. Rech. Opér. Sér. Rouge **3** (1969) 35–43.
- [32] B.T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comput. Math. Math. Phys. **9** (1969) 94–112.
- [33] L.I. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992) 259–268.
- [34] A. Sahiner, N. Yilmaz, *Double stranded smoothing technique in energy minimization problems*, Numer. Algebra Control Optim. **14** (2024) 80–88.
- [35] S. Voronin, G. Ozkaya, D. Yoshida, *Convolution based smooth approximations to the absolute value function with application to non-smooth regularization*, arXiv preprint arXiv:1408.6795 (2014).
- [36] W. Wang, Y. Chen, *An accelerated smoothing gradient method for nonconvex nonsmooth minimization in image processing*, J. Sci. Comput. **90** (2022) 31.
- [37] X. Wang, G. Yuan, L. Pang, *A class of new three-term descent conjugate gradient algorithms for large-scale unconstrained optimization and applications to image restoration problems*, Numer. Algorithms **93** (2023) 949–970.
- [38] C. Wu, J. Zhan, Y. Lu, J.S. Chen, *Signal reconstruction by conjugate gradient algorithm based on smoothing ℓ_1 -norm*, Calcolo **56** (2019) 42.

- [39] C. Wu, J. Wang, J.H. Alcantara, J.S. Chen, *Smoothing strategy along with conjugate gradient algorithm for signal reconstruction*, J. Sci. Comput. **87** (2021) 1–18.
- [40] N. Yilmaz, A. Sahiner, *New smoothing approximations to piecewise smooth functions and applications*, Numer. Funct. Anal. Optim. **40** (2019) 513–534.
- [41] N. Yilmaz, A. Sahiner, *On a new smoothing technique for non-smooth, non-convex optimization*, Numer. Algebra Control Optim. **10** (2020) 317–330.
- [42] N. Yilmaz, A. Sahiner, *Generalization of hyperbolic smoothing approach for non-smooth and non-Lipschitz functions*, J. Ind. Manag. Optim. **18** (2022) 1111–1125.
- [43] N. Yilmaz, A. Kayacan, *A new smoothing algorithm to solve a system of nonlinear inequalities*, Fundam. J. Math. Appl. **6** (2023) 137–146.
- [44] N. Yilmaz, A. Sahiner, *Smoothing techniques in solving non-Lipschitz absolute value equations*, Int. J. Comput. Math. **100** (2023) 867–879.
- [45] N. Yilmaz, *Introducing three new smoothing functions: Analysis on smoothing-Newton algorithms*, J. Math. Model. **12** (2024) 463–479.
- [46] N. Yilmaz, A. Zeytinoglu, *A novel class of exact penalty function approach for optimization problems with inequality constraints*, J. Math. Model. **13** (2025) 153–167.
- [47] W. Yin, S. Osher, D. Goldfarb, J. Darbon, *Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing*, SIAM J. Imaging Sci. **1** (2008) 143–168.
- [48] C. Zeng, C. Wu, R. Jia, *Non-Lipschitz models for image restoration with impulse noise removal*, SIAM J. Imaging Sci. **12** (2019) 420–458.
- [49] H. Zhu, Y. Xiao, S.Y. Wu, *Large sparse signal recovery by conjugate gradient algorithm based on smoothing technique*, Comput. Math. Appl. **66** (2013) 24–32.